**VIETNAM NATIONAL UNIVERSITY, HO CHI MINH CITY**

**UNIVERSITY OF INFORMATION TECHNOLOGY**

**FACULTY OF INFORMATION SYSTEMS**

**\*\*\*\*\*\***

****

**DECISION SUPPORT & BUSINESS INTELLIGENCE APPLICATIONS**

**DATA MINING**

**Lecturer:**MSc. Nguyen Thi Kim Phung

**Class:** MSIS4263.P21.CTTT

**Student:** Ngo Manh Khuong - 21522241

Ho Chi Minh city, May 2025

**Conclusion from the Visualizations**

A diagram of a graph

AI-generated content may be incorrect.

A graph of different colored bars

AI-generated content may be incorrect.

1. **Key Feature Differences Between Correct and Incorrect Predictions:**
   * True Positives (TP) (correctly predicted diabetics) tend to have higher values of:
     + Glucose
     + Insulin
     + BMI
     + Age
   * These features appear to be strong indicators of diabetes, aligning with medical knowledge.
   * False Negatives (FN) (actual diabetics incorrectly predicted as non-diabetics) generally have lower Glucose and Insulin levels than TPs but still higher than True Negatives (TNs), suggesting borderline values might lead to misclassification.
2. **False Positives (FP) vs True Negatives (TN):**
   * False Positives show elevated Glucose and Insulin compared to TNs, though not as high as TPs, implying the model might be overly sensitive to these features, sometimes flagging non-diabetic individuals as diabetic.
3. **Radar Chart Patterns:**
   * The radar chart shows that Glucose and Insulin create the largest separations between categories, indicating they are the most influential features in the model.
   * Pregnancies, Skin Thickness, and Diabetes Pedigree Function contribute less variance across prediction categories.

**Rule 1: Predict Positive (Diabetic)**

If:

* Glucose > 130 AND
* Insulin > 100 AND
* BMI > 35 AND
* Age > 35
* Predict Outcome = 1 (Diabetic)

**Rule 2: Predict Negative (Non-Diabetic)**

If:

* Glucose < 110 AND
* Insulin < 80 AND
* BMI < 30 AND
* Age < 35
* Predict Outcome = 0 (Non-Diabetic)

**Models:**

1. **Autoencoder (77.08%)**
2. XGBoost (76.04%)
3. CNN (75.52%)

A graph of different colored bars

AI-generated content may be incorrect.

**END**